Determinants of Water Demand in Marginalized Urban Areas in Lima-Peru and their Importance in Improving Social Welfare

By

HUAMANI ANTONIO, Sandro Alejandro

THESIS

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC POLICY

2019

Determinants of Water Demand in Marginalized Urban Areas in Lima-Peru and their Importance in Improving Social Welfare

By

HUAMANI ANTONIO, Sandro Alejandro

THESIS

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC POLICY

2019

Professor KIM, Taejong

Determinants of Water Demand in Marginalized Urban Areas in Lima-Peru and their Importance in Improving Social Welfare

By

HUAMANI ANTONIO, Sandro Alejandro

THESIS

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC POLICY

Committee in charge:

Professor KIM, Taejong, Supervisor ______________________

Professor Nam, Il-Chong ______________________

Approval as of May, 2019

ABSTRACT

DETERMINANTS OF WATER DEMAND IN MARGINALIZED URBAN AREAS IN LIMA-PERU AND THEIR IMPORTANCE IN IMPROVING SOCIAL WELFARE

By

Huamaní Antonio, Sandro Alejandro

This research studies part of the water market from marginalized urban areas in Lima-Peru, where households are not provided potable water service with pipes by the Lima water company. Given this, consumers of marginalized areas pay higher prices from tank trucks services than consumers of the Lima water company. Therefore, the main purposes of this study are: (1) to identify the determinants of water demand using an instrumental variable regression, and (2) to know how social welfare changes when water price varies. For measuring this change, I use concepts such as consumer surplus, producer surplus, equivalent variation and compensating variation. I use data from the survey conducted by the Direct Marketing Research and Consulting, and the National Superintendence of Sanitation Services of Peru (SUNASS). Finally, the results suggest that consumption of water depends significantly on water price, number of people per household, the storage capacity of water that each household has and the level of income. Likewise, we found out that there is a strong negative relation between water price and social welfare; thus, public policies to regulate the level of water price are strongly recommended.

JEL Codes: D12, D60, I31.

Keywords: Lima-Peru, water service demand, social welfare.

ACKNOWLEDGEMENTS

First of all, I am grateful to my advisers Professor Taejong Kim and Professor Il-Chong Nam for their extensive insights, feedbacks, support, and encouragement throughout this dissertation. Their guidance has substantially improved this work and it has helped me to develop a topic related to the water sector. I also thank my KDI fellows, especially who are from Latin America, for the sleepless nights and for all their friendship and company during my stay in Korea. On the other hand, I must express my gratitude to my parents, Vicente Huamaní and Cristina Antonio, my brothers, Jhosel, Kevin and Miguel, and my sister Marilyn; likewise, my sincere gratitude to my fiancé Mayra Pizarro and her family, this achievement would not have been possible at all without their scolding, counsel, love, sacrifice, trust, and support. Finally, I would like to dedicate this work to my future princess, Brianna Avril Huamaní Pizarro, who was my main motivation to complete the master degree and finish this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

1.1. Background

This research is focused on the water market in marginalized urban areas in Lima-Peru, where consumers are households who do not have access to potable water service with pipes from the Lima water company, SEDAPAL[1]. Given this, the water suppliers in marginalized areas are small private companies that provide water via tank trucks. There is no regulation in this market or any kind of supervision by the potable water regulator. Using data from the Direct Marketing Research and Consulting (2015), I describe some characteristics about this water market.

Regarding to demand, it is important to mention that households, who do not have access to drinking water service, are around 10% out of total households of Lima (SUNASS, 2015), about 195,600 households; and 90% out of them live in marginalized urban areas due to their economic condition. Likewise, 86% out of total consumers buy water from tank trucks, 11% from a neighbor and 3% from others. Since they do not have water tab in their house, they store water using different containers; thus 59% out of total demanders use cylinders or drums, 31% use water tanks and 10% use water bucket and tub (Huamani, 2017).

One relevant characteristic in this market is that some households transport the water from the purchase point to their house (Bonifaz & Aragón, 2008). Overall, 18% out of total consumers transport water around one block, 6% around two blocks and 7% more than two blocks. Also, 49% out of households who transport water, the father or mother are in charge of this activity, 14% out of these households spend around 10 minutes doing this activity, 18% spend between 10 and 20 minutes, 37% spend between 20 and 30 minutes, and 31% spend more than 30 minutes.

[1] "*Servicio de Agua Potable y Alcantarillado de Lima*" in Spanish.

Another characteristic of this market is the intensive reuse of water due to scarcity. In this regard, the water that is most reused is for laundry, then from dish washing, clothes washing, hand washing, showering, and food preparation (Huamani, 2017).

Regarding supply, in most cases there is no competition between water suppliers because 58% out of total households buy water from one supplier, meanwhile, 16% from two suppliers and 26% from more than two suppliers. However, 42% of households, who buy water from more than one supplier, consider that there is no competition among their suppliers. Likewise, these households consider that competition factors are frequency of water supplying per week, water quality and water price level, in this order of importance.

Finally, I use the Direct Marketing Research and Consulting survey, after eliminating some outliers, to estimate the average water price of this market which is around 15.7 PEN[2]/m^3, and the average monthly consumption of water which is 6 m^3. For this level of water consumption, SEDAPAL is charging an average price below 2 PEN/m^3 to its users, and given that the lack of intensive competition, we can infer that this market would have similar characteristics to no-regulated monopoly or duopoly market with an inefficient allocation of resources that will be reduced if there is a reduction in the water price (Harberger, 1954).

1.2. Importance and objectives of this study

The primary objective of this study is to identify the determinants of water demand in marginalized urban areas in Lima-Peru for households who do not have access potable water with pipes from SEDAPAL. I estimate the water demand equation using an Instrumental Variable approach regression due to reverse causality problem between the water consumption and the water price. Likewise, the secondary objective of this study

[2] Currency of Peru, "Peruvian Sol".

is to know how social welfare changes with water price variations. To measure these changes, I define social welfare as the sum of welfare consumer and welfare producer.

Knowing what the determinants of water demand are and how variation in water price can improve the social welfare is very important for designing effective public policies which will allow poor households from Lima to maximize the gain in welfare. Likewise, this study is relevant because there is a lack of studies on the characteristics of water market in marginalized urban areas. Findings of this study can be useful for public institutions such as SUNASS[3], the drinking-water regulator, and the Ministry of Construction and Sanitation of Peru.

1.3. Research question and hypotheses

The main research question of the study is: what are the determinants of water demand in marginalized urban areas in Lima-Peru for households who do not have access to potable water service with pipes? The possible answer is my first hypothesis: the variation of water demand for households, who do not have access to drinking water service in Lima – Peru, is mainly explained by the variation of the following explanatory variables: i) water price, ii) family income, iii) water transportation time, iv) size of family, v) storage capacity of water.

The second research question of this study is: how does social welfare change when the water price changes? The possible answer is my second hypothesis: reducing water price can significantly increase social welfare because an increase in consumer welfare is greater than a decrease in producer welfare.

[3] "*Superintendencia Nacional de Servicios de Saneamiento*" in Spanish.

II. LITERATURE REVIEW

2.1. Classical Demand Theory

In this section, I present a summary of the Classical Demand Theory which is considered a preference-based approach of consumer demand to know which variables determine the demand. This theory mentions that the demand is formed by the union of optimal points where the consumer maximizes his utility subject to a budget constraint.

The analysis of consumer behavior begins by specifying the consumer's preferences ($\succsim$) over the commodity bundles (X). This preference can define to rational consumers if it satisfies two assumptions: complete[4] and transitive[5] (Kreps, 2012).

Likewise, there are two additional assumptions called desirability and convexity.[6] The first one refers that there is more preferable to have large amounts of goods than small (monotonicity). The second one concerns the trade-off that consumer is willing to make among different goods. It can be interpreted in terms of "diminishing marginal rates of substitution" (Mas-Colell, Whinston & Green, 1995). Both assumptions help us representing preferences using utility functions *U(X)* which are increasing and quasiconcave, but not continuous. For the existence of continuous *U(X)* is necessary to have one more assumption called continuity (Kreps, 2012).

Then, having a rational consumer who has a continuous utility function, we can explore the following utility maximization problem:

$$Max\ U(X)\ s.t\ pX = I, \tag{1}$$

where p is a vector of prices and I is the level of consumer's income. For simplicity, we shall suppose that the choice of consumer is considering just two goods (x_1 and x_2), so the previous constrained-optimization problem is replaced by:

---

[4] For all $x_1, x_2 \in X$, we have $x_1 \succsim x_2$ or $x_2 \succsim x_1$ or both.

[5] For all $x_1, x_2, x_3 \in X$, we have $x_1 \succsim x_2$ and $x_2 \succsim x_3$, then $x_1 \succsim x_3$.

[6] If $x_2 \succsim x_1$ and $x_3 \succsim x_1$, then $\alpha\, x_2 + (1-\alpha)\, x_3 \succsim x_1$ for $\alpha \in [0,1]$

$$Max\ U(x_1, x_2)$$

s.t

$$p_1x_1 + p_2x_2 = I \tag{2}$$

For getting the first-order condition, first we must write the Lagrangian function of this optimization as follows:

$$L = U(x_1, x_2) + \lambda(I - p_1x_1 - p_2x_2) \tag{3}$$

Then, the first-order condition is:

$$\frac{\partial L}{\partial x_i} = \frac{\partial U(x_1, x_2)}{\partial x_i} - \lambda p_i = 0 \tag{4}$$

$$\frac{\partial L}{\partial \lambda} = I - p_1x_1 - p_2x_2 = 0 \tag{5}$$

It would not be necessary to satisfy the second-order condition because we have already known that it is a maximization. Then, solving these simultaneous equations, we can find the following Walrasian or ordinary demands (Chiang, 2005).

$$x_1^* = x_1(p, I)$$

$$x_2^* = x_2(p, I) \tag{6}$$

As we can see, the demand depends on a vector of prices, *p*, and the income, *I*, however, the demand can also depend on other economic and social variables (Varian, 1992). It is important to mention that Walrasian demands must satisfy properties such as homogeneity of degree zero in (*p*, *I*), Walras' law, convexity, and uniqueness (Mas-Colell, Whinston & Green, 1995). Additionally, if x_1^* and x_2^* are introduced in objective function *U(X)*, we can get the indirect utility function which is the maximum utility that a consumer can obtain given *p* and *I*:

$$V(x_1^*, x_2^*) = V(p, I) \tag{7}$$

Finally, if there are *n* consumers with rational preference relations, there could be an aggregate demand which can be written as:

$$x(p, I_1, I_2, .., I_n) = \sum_{i=1}^{n} x_i(p, I_i) \quad (8)$$

The aggregate demand can be defined as the sum of the individual consumer demand, where this aggregation introduces three additional determinants of demand such as the number of consumers, the distribution of preferences among the consumers, and distribution of incomes among consumers (Jehle & Reny, 2011).

2.2. Empirical studies related to water demand

It is important to mention that almost all empirical studies which present a water demand estimation have used a sample of households who have access to drinking water demand from water companies, where consumers face increasing-block prices, i.e., nonlinear price (Jaramillo-Mosqueir, 2005). However, these cases are slightly different from the case addressed in this research, because, as we mentioned, our sample is composed by households who do not have access to potable water with pipes, so they have to buy the water from tank trucks and pay a linear water price.

Respect to the methodology, the authors initially discussed about whether the price variable in a demand equation should be marginal price or the average price (Billings and Agthe, 1980). However, after Taylor (1975) and Nordin (1976), studies have focused on differences instead of marginal or average price to account for the lump sum transfers implied by increasing-block price. Likewise, we can find studies such as Opaluch (1984), which have considered marginal price and takes the difference inside the specification. In this context, authors have used several econometric techniques among OLS, 2SLS, instrumental variables and likelihood maximum estimation (Hewitt and Haneman, 1995). In recent studies such as Hewitt and Haneman (1995), Olmstead (2009) or Miyawaki (2011), they have used a the discrete continuous choice model proposed by Hausman (1985) which is based in nonlinear budget sets due to increasing-block price.

Moreover, there are abundant studies for developed countries especially in urban areas where households have access to potable water services with pipes. For instance, Forster and Beattie (1979), Billings and Agthe (1980), Al-Qunaibet and Johnston (1985), Hewitt and Hanemann (1995), Pint (1999), Nauges and Thomas (2000), Cavanagh, Hanemann and Stavins (2001), Olmstead, Hanemann and Stavins (2007), Miyawaki, Omori and Hibiki (2011), and among others.

For developing countries, the literature is small. In particular, for Latin America, we can find principally studies written by Jaramillo-Mosqueir (2005) for Mexico, and Jimenez, Orrego, Vasquez and Ponce (2017) for Manizales city in Colombia, both using discrete continuous Choice model. Finally, for Peru, Ortiz and Bendezu (2006) used a three-stage least squares (3SLS) method; likewise, Huamani (2017) shows a preliminary estimation of the demand for household who do not have access to potable water with pipes using a OLS estimation.

Finally, respect to the set of variables used in these estimations, most of the studies have considered the water price, family income, house size (number of rooms, bathrooms, etc.) and family size as principal explanatory variables. Additionally, Huamani (2017) considered other variables pertaining to the situation where households do not have access to potable water with pipes such as storage capacity of water in tanks and frequency that the tank trucks provide the service.

2.3. Social welfare

The discussion on how to specify the social welfare function (*W)* is limited. However, even though there are same certain general restrictions which eliminate the possibility of having an ideal relationship between individual preferences and the social ordering (Arrow, 1951), there is a favorite assumption which states that the social welfare function is individualistic and provides an index which aggregate the individual utilities of economic agents, $W(u_1(x), \dots, u_r(x))$, assuming that W increases with any u_i given the utilities of all others (Sen, 1997).

In this context, the most widely used approach is known as utilitarianism which considers that the sum of the individual utilities is a measure of social welfare of the form: $W = \sum_{i=1}^{r} u_i$. However, the problem with this approach is that maximizing the sum of individual utilities is unconcerned with the interpersonal distribution of the sum, i.e., it does not matter egalitarian reasons (Sen, 1997). An alternative approach with much stronger egalitarian criteria has been proposed by Rawls (1985), where the social welfare is defined by the function: $W = \min(u_1, \dots, u_r)$. Then, the social welfare of an allocation depends only on the welfare of the worst off agent (Varian, 2014).

On the other hand, according to Microeconomic theory, the total welfare in a market is the sum of consumer surplus (CS) and producer surplus (PS). Therefore, changing of exogenous economic variables such as price of the good or income generates variation in the social welfare among variations in consumer surplus and producer surplus (Stiglitz, 2000). I explain in detail this concept in the methodology section.

Additionally, to measure the variation in consumer welfare, we also can resort to the normative side of consumer theory, which is called the welfare analysis (Mas-Colell, Whinston & Green, 1995). First, we should construct a money metric indirect utility for any price vector p. However, depending if we use the initial price, p^0, or a new price, p^1, we can obtain two consumer well-known measures (Hicks, 1939), called: (i) Equivalent variation (EV) and (ii) compensating variation (CV).

The EV can be defined as the money amount that the consumer is willing to accept for facing a price change. In other words, the change in his wealth that is equivalent to price change in terms of welfare impact at p^0. On the other hand, the CV can be defined as the money amount that planner must compensate the consumer for the price change after it occurs to keep the original utility level at p^1. Formally, we can define them as follows:

$$EV(p^0, p^1, I) = e\big(p^0, V^1(p^1, I)\big) - e\big(p^0, V^0(p^0, I)\big) = e(p^0, V^1) - I \qquad (9)$$

$$CV(p^0, p^1, I) = e(p^1, V^1(p^1, I)) - e\big(p^1, V^0(p^0, I)\big) = I - e(p^1, V^0) \qquad (10)$$

where $e(p, V)$ is the expenditure function which depends on prices and indirect utility. Additionally, the EV and CV have interesting representations in term of Hicksian demand curve, $x_h(p, V)$, which is a result of the expenditure minimization problem (Kreps, 2012). then, if we define $I = e(p^0, V^0) = e(p^1, V^1)$, the EV can be written as:

$$EV(p^0, p^1, I) = e(p^0, V^1) - e(p^1, V^1) = \int_{p_1^1}^{p_1^0} x_{h1}(p_1, p_2, V^1) dp_1 \qquad (11)$$

Similarly, the CV can be written as:

$$CV(p^0, p^1, I) = e(p^0, V^0) - e(p^1, V^0) = \int_{p_1^1}^{p_1^0} x_{h1}(p_1, p_2, V^0) dp_1 \qquad (12)$$

Finally, for a normal good we should expect the following relationship for a decreasing of price, p_1 (Mas-Colell, Whinston & Green, 1995):

$$EV(p^0, p^1, I) > \Delta CS > CV(p^0, p^1, I) \qquad (13)$$

2.4. Water service and social welfare

According to some studies, there would be a relationship between water service and social welfare because providing water service for poor people can increase the social welfare and generates social benefits that will be greater than social costs of implementing this public policy. For example, Hutton, Haller and Bartran (2007) have mentioned that there are at least two benefits of access to drinking water and sewerage. First, people can avoid several diseases as diarrhea or cholera and save money in buying medicines. Second, people can avoid spending time of moving water from the sale point to their house.

In the Peruvian case, Bonifaz and Aragon (2008) stated that people who do not have access to potable water with pipes pay higher price for buying water in tank trucks. Therefore, if they are provided with pipe water service by SEDAPAL consumers can pay a lower price and have a higher water consumption. Likewise, Bonifaz and Aragon (2008) estimated a cost overrun for households due to lack of water access in USD 160.23 million based on transaction cost theory by Ronald Coase.

Moreover, Oblitas (2010) identified three large areas where sanitation services contribute to improve the social welfare: (1) reduction of poverty levels, (2) increasing attendance levels and school performance, and (3) improving living condition. Oblitas (2010) mentioned that sanitation service reduces poverty because it decreases the expenses of medical attention, medicines and time of care for the sick; likewise, it decreases mortality rates and the family expenditure of water.

In the same way, Huamani (2017), using a survey of 1000 households from Lima-Peru, quantified three more benefits which are: (1) money saving, (2) health benefits, and (3) time saving. They were estimated in USD 26 million, USD 4 million and USD 56 million per year, respectively. It means that the annual social benefit for providing drinking water to households who do not have access to is around USD 86 million. On the other hand, the infrastructure investments necessary for providing drinking water was estimated in approximately USD 730 million. Additionally, the operating expense was estimated in USD 27 million per year. Finally, using a cash flow of 30 years, Huamani (2017) demonstrated that social benefits are greater than social costs; thus, social welfare is bigger when potable water is provided.

III. METHODOLOGY

3.1. Information collected

For estimating the water demand, I use the data collected from the survey conducted by the Direct Marketing Research and Consulting in 2015. This was requested by SUNASS, the drinking-water regulator of Peru, to know how poor families obtain water, how much they consume, how much they pay per cubic meter, as well as the socioeconomic characteristics of consumers in Metropolitan Lima. After checking the information, I selected a sample size of 604 households because some observations presented atypical behavior.

Based on the SEDAPAL Tariff Study for the 2015-2020 regulatory period and prepared by SUNASS, the population of households who do not have access to water services is around 195,600, which was estimated using the information on: number of people in Lima (9.6 million), density (4.1 people per house) and drinking water coverage (91.6%). From the aforementioned information, a sampling error of 3.98% was estimated, which shows the reliability of the information to make statistical inferences.

Table 1. Sampling Error

Variable	**Value**
Population (Household)	195 600
Sample	604
Expected frequency of the parameter	0.5
Z(α=0.05)=1.96	1.96
Sampling Error	3.98%

It is important to mention that the sample is probabilistic, multistage, and systematic. Each household had a known probability of being selected, which means that the choice of observations was completely random. Likewise, the selection of observations was by three stages: selection by population centers, selection by conglomerates, and selection by households, respectively. Finally, it was systematic sampling because after making groups of households, observations were selected from each group.

3.2. Estimation of water demand using an Instrumental Variable regression

3.2.1. Why should we use Instrumental Variables (IV)?

The Instrumental Variable (IV) regression approach can be used to solve endogeneity problem of one or more explanatory variables. This problem can be associated to mainly three cases which are: (i) omitted variable, (ii) measurement error and (iii) reverse causality or simultaneity. The endogeneity problem leads a biased and inconsistent OLS estimator (Wooldridge, 2016), so it cannot longer give a causal interpretation or be the marginal effect on the dependent variable of an exogenous change in an explanatory variable.

In this study, there is reverse causality problem. Even though the consumed water quantity depends on the water price, the water price is also determined by consumption of water from household. As mentioned before, households that do not have access to drinking water service obtain the water from tank trucks which charge a water price in function of the type of container (water tank, reservoir, cylinder, tub, etc.) used for buying. It is easy to infer that these containers can store different volumes of water by liter, and households use them depending on the quantity of water they want to consume. Then, the water price is determined simultaneously or jointly by water consumption. In conclusion, the water price could be endogenous and would be correlated with the error term.

In this regard, it is important to define an instrument to correct the reverse causality problem and get unbiased and consistent estimators: however, this goal will be achieved only if we find valid instruments (Cameron & Trivedi, 2009) that satisfy two conditions: (1) relevance and (2) exogeneity.

To clearly understand these conditions, let's define the following multiple regression model:

$$y_i = \beta_0 + \beta_1 p_i + \beta_2 x_{2i} + u_i \tag{14}$$

where y is the dependent variable (water demand or consumption of water), p (price water) and x_2 are explanatory variables and u is the error term. Let us suppose that p is determined simultaneously by y, so p is endogenous, i.e., : $cov(p, u) \neq 0$. Therefore, we should introduce an instrument, called z, which satisfy the conditions mentioned:

- Relevance: This condition demands that z should have a causal effect on p, so $cov(z, p) \neq 0$. Likewise, changes in z can be associated with changes in y, but only through p.
- Exogeneity condition: This condition demands that z should be exogenous, so it should not be correlated with u, $cov(z, u) = 0$. Likewise, z does not share common causes with y.

Therefore, one of the challenges in this study is to find instruments which must have correlation with the water price, affect the consumption of water via price and be exogenous.

3.2.2. Two Stage Least Square (2SLS)

Two-Stage Least Squares (2SLS) regression is an extension of the OLS method and it can be used in the analysis of equations. In general, this technique, to obtain the best IV estimator, has a two-stage procedure:

i) In the first stage, we regress the endogenous regressor, *p*, on all exogenous variables, x_2, and the instrumental variable(s), z.

$$p_i = \pi_0 + \pi_1 z_i + \pi_2 x_{2i} + v_i \quad (15)$$

ii) In the second stage, we regress the dependent variable, *y*, on all exogenous regressors, x_2, and the prediction of *p*.

$$y_i = \beta_0 + \beta_1 \hat{p}_i + \beta_2 x_{2i} + u_i \quad (16)$$

In summary, the 2SLS regression breaks *p* into two parts, one that can be correlated with *u*, and the other is not. Then, by deleting the part which is correlated with the error term, 2SLS can provide unbiased and consistent estimator of β_1.

In this regard, one important aspect to use 2SLS is to know the relationship between the number of instruments, *dim(z)*, and the number of endogenous regressors, *dim(x)*. This relation can be of three different types: (i) just-identified case if *dim(z)* = *dim(x)*, (ii) under-identified case if *dim(z)* < *dim(x),* and (iii) over-identified if *dim(z)* > *dim(x)*.

In the first case, transforming the zero correlation condition, $E\{z'_i(y_i - x'_i\beta)\} = 0$, in a vector form, $Z^{'}(y - X\beta) = 0$, and solving, we can get the IV estimator:

$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'y \quad (17)$$

In the second case, there will be an inconsistent IV estimator, so it is recommended obtaining enough instruments, even just one in applications with a single endogenous regressors. That is a challenge because it can require considerable ingenuity or access to unusually rich data (Cameron & Trivedi, 2009).

Finally, in the third case, $Z^{'}(y - X\beta) = 0$ cannot be a solution because there is a system of *dim(z)* equations. Therefore, the more-efficient estimators can be obtained using the Two-Stage Least Squares (2SLS) estimator as follows:

$$\hat{\beta}_{2SLS} = \{X'Z(Z'Z)^{-1}Z'X\}^{-1}X'Z(Z'Z)^{-1}Z'y \quad (18)$$

3.2.3. Diagnostics for weak instruments

Using a weak instrument can lead to substantial problems. For example, if there is even a small correlation between the instrument and error term (weak exogeneity) and/or the instrument has just a little correlation with the explanatory variable (weak relevance), the estimates would be inconsistent and biased (Wooldridge, 2016). Therefore, it is very important to detect weak instruments and avoid them.

In this regard, Staiger and Stock (1997), and Stock and Yogo (2005) proposed a method to detect weak instruments based on the size of the t statistic or F statistic where there is one instrument or are more than two instruments, respectively.

Thus, in the case of 2SLS, it is necessary to focus on the first-stage F statistic for excluded instruments, and then test a hypothesis in which the null hypothesis is that the instruments are weak and the alternative is that they are strong. For ensuring that the null hypothesis is rejected, the authors proposed that F statistic should be larger than 10. We should use the same rule using t statistic when we have just one instrument.

3.3. Measure of the social welfare

For finding a measure of the social welfare, first, we are going to suppose that the structure market is similar to no-regulated monopoly, because, as we mentioned in the background section, most of the consumers of water face just one supplier. figure 1 shows the current situation of water market in marginalized urban area in Lima-Peru., where we have a current water price (p^0) which would be above the equilibrium price and current consumption of water (y^0) which would be below the equilibrium quantity. In consequence, this situation leads to a deadweight loss (DL) due to allocative inefficiency (Harberger, 1954).

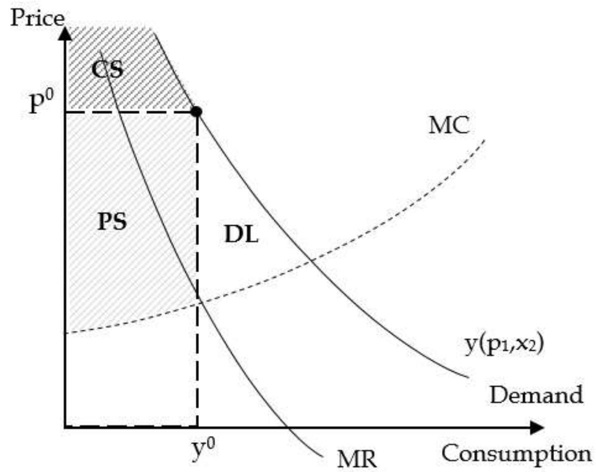

Figure 1. Current Situation of Water Market

Based on the aforementioned, I consider a measure of social welfare as the sum of consumer welfare and producer welfare. That means, initially, the sum of all consumer (n) surpluses and all producer (m) surpluses, as an approximation of Utilitarian function.

$$W(p_1, x_2) = \sum_{i=1}^{n} CS_i\big(y(p, x_2)\big) + \sum_{j=1}^{m} PS_j\big(y(p)\big) \tag{19}$$

where, consumer surplus (CS) is measured by the consumer's willingness to pay minus the price actually paid, and producer surplus (PS) is measured by price received minus the cost of production.

3.3.1. Variations in social welfare

Previously, we notice that the social welfare depends on water price, *p*, and other explanatory variables, x_2, so variations in *p* and x_2 lead to variations in the social welfare. Likewise, variations in *p* and x_2 can change water demand curve, *y(.)*, in different ways and magnitudes. However, in this study, we will just concentrate in variation in water price, because it is the determinant which can be regulated in short term through public policies.

A variation in the water price, with x_2 fixed, causes the consumer to choose a different market basket. In this study, I assume that water price can decrease until having equilibrium price (p^1) and equilibrium quantity (y^1) as shown in figure 2.

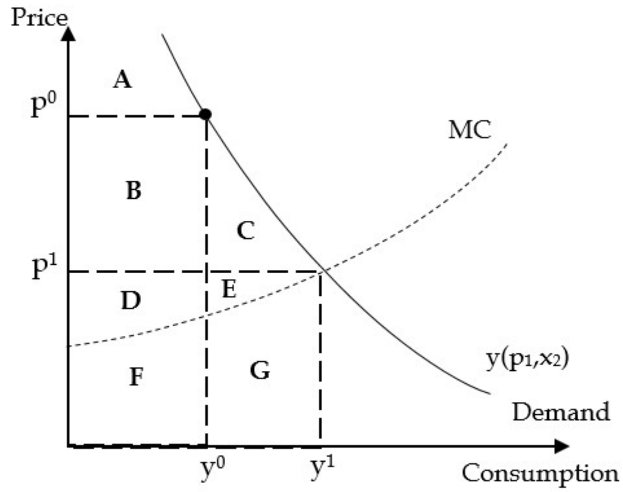

Figure 2. Variation in Price

Knowing the variation in social welfare is important to identify the level of consumer surplus and producer surplus before and after water price changes. According to before-after analysis, we found out that the variation in social welfare is area C+E, as shown in table 2.

Table 2. Variation in Social Welfare (Before - After)

Variable	Before	After	Variation
Consumer surplus	A	A+B+C	B+C
producer surplus	B+D	D+E	E-B
Social welfare	A+B+D	A+B+C+D+E	**C+E**

Mathematically, to calculate C+E area, it is necessary to estimate the variation on consumer surplus (ΔCS), which is the area B + C:

$$B + C = \int_{p1}^{p0} y(p, x_2) dp \tag{20}$$

Then, it is also necessary to estimate variation on producer surplus (ΔPS), finding the area E – B:

$$E - B = p^1(y^1 - y^0) - \int_{y^0}^{y^1} MC(y)\, d_y - (p^0 - p^1)y^0$$

$$E - B = p^1(y^1 - y^0) - AC(y^1 - y^0) = (p^1 - AC)(y^1 - y^0) - (p^0 - p^1)y^0 \tag{21}$$

Finally, the variation on social welfare in this case (Δp) is defined by:

$$\Delta W = \int_{p1}^{p0} y(p, x_2) dp + (p^1 - AC)(y^1 - y^0) - (p^0 - p^1)y^0 \tag{22}$$

Likewise, as reference, Appendix A shows how the social welfare would change when there are variations in x_2, with p fixed.

Additionally, instead of variation on consumer surplus we can use other measure of consumer welfare such as compensating variation (CV) or equivalent variation (EV) as it was mentioned previously in the literature review. Figure 3 shows the CV and EV, in term of Hicksian demand curve, for a reducing of water price, p_1.

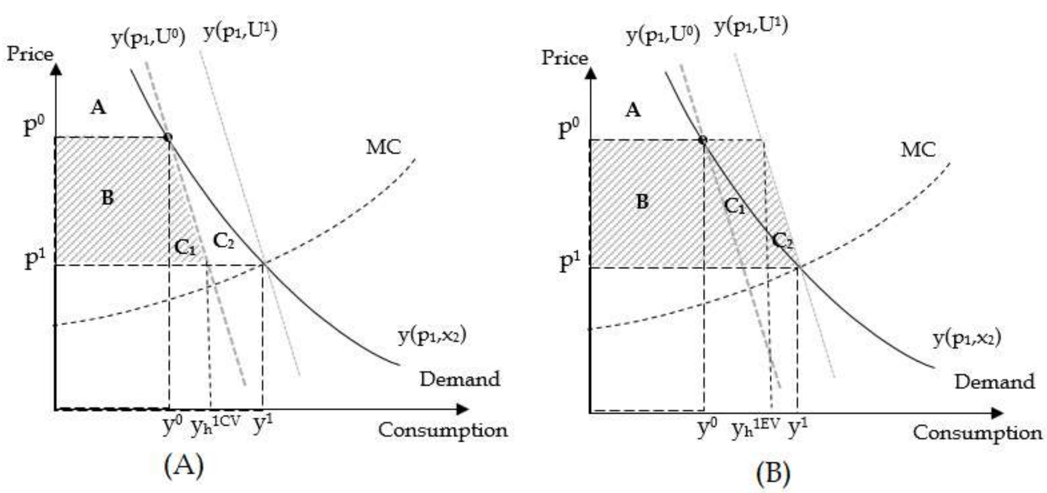

Figure 3. Equivalent and Compensating Variation

In the part A of figure 3, the area B + C_1 represents the value of VC, which should be smaller than the variation of consumer surplus showed (B + C) in the figure 2 because of C > C_1. Likewise, in the part B, the area B+ C_1+C_2 represents the value of EV which should be larger than the variation of consumer surplus because of C < C_1 + C_2. Then, using the previous figure, we can measure the CV as follows:

$$B + C_1 = (p^0 - p^1)y^0 + (p^0 - p^1)(y_h^{1C} \quad - y^0)/2 \tag{23}$$

Likewise, we can measure the EV as follows:

$$B + C_1 + C_2 = (p^0 - p^1)y_h^{1E} \quad + (p^0 - p^1)(y^1 - y_h^{1EV})/2 \tag{24}$$

Finally, the values of y_h^{1CV} and y_h^{1EV} are showed in Appendix B.

IV. RESULTS AND DISCUSSION

4.1. Estimation of the water demand

As mentioned in the methodology, I estimate the water demand through Two-Stage Least Squares (2SLS) using instrumental variables and a cross-sectional database. To do this, I define first the model for estimation, then I choose the valid instrumental variables. Finally, I do a specification analysis about the 2SLS estimation. Therefore, my estimating model is:

$$\begin{aligned} LnQ_i = \beta_0 + \beta_1 LnP_i + \beta_2 Number_{members,i} + \beta_3 Transportation_{time,i} \\ + \beta_4 StorageCap_i + \beta_5 DummyIncome_i + u_i \end{aligned} \tag{25}$$

In equation (25), we have the logarithm of monthly consumption of water in cubic meter (m^3) per household (LnQ), the logarithm of water price (PEN/m^3) per household (LnP), number of people per household (Number_members), how many minutes each households spends to transport the water from purchase point to house (Transportation_time), the storage capacity of water in liter that each household has (StorageCap), finally, a dummy variable where it takes the value of 1 if the income is bigger than PEN 315 monthly and 0 otherwise (DummyIncome).

Table 3 presents the summary statistics of the main variables:

Table 3. Descriptive Statistics for Dataset

Variable	Units	Mean	Standard Deviation	Minimum	Maximum
Q (Monthly consumption)	m^3	6.04	3.94	0.22	26.4
P (Water price)	PEN/m^3	15.75	6.04	0.90	50.00
Number_members	number	4.76	1.81	1.00	12.00
Transportation_time	minute	3.30	8.65	0.00	30.00
StorageCap	liters	0.64	0.47	0.02	2.60
DummyIncome	dummy	0.31	0.462	0.00	1.00

As mentioned in the methodology section, it is necessary to use instrumental variables because water price is determined simultaneously or jointly by consumption of water. Therefore, the water price could be endogenous and would be correlated with the error term. Thus, in the following sub-section, I define valid instrumental variables.

4.1.1 Choice of instrumental variables

The variables used as instrumental variables are the set of the dummy variable by each district, where households without water service are located, due to the relevance and exogeneity condition. As a reference for checking the relevance condition, I see whether districts have a causal effect on water price, $cov(district, p) \neq 0$, through the figure below, which shows a certain relationship between average price and districts.

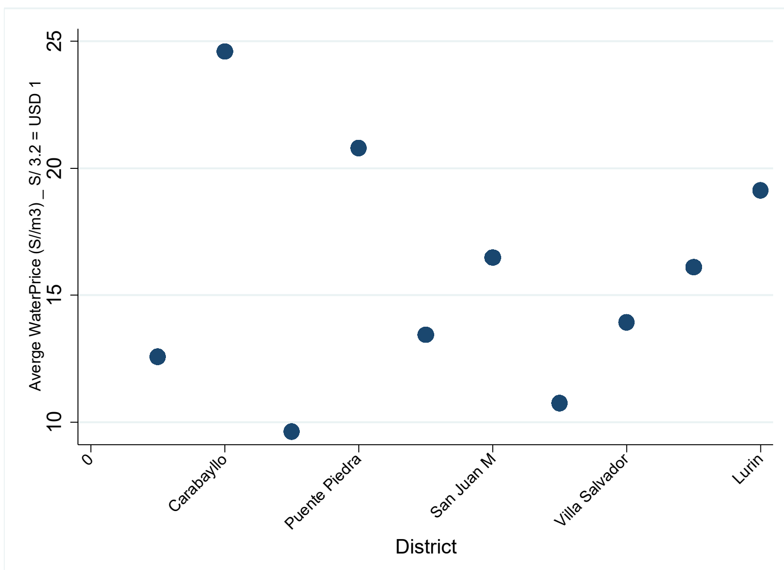

Figure 4. Average Water Price by District

We can deduce that the mentioned relationship exists because the competition between supplier is different in each district (it could be explained for the difference of coverage of drinking-water service) and this situation can lead to difference in average water price. Districts with more competition would have a lower water price. This statement is consistent due to the correlation of -0.38 between the level of competition and district.

To make sure that there is a correlation between district and water price, I checked the result of the first stage of 2SLS. It shows that the p-value of most of the districts is lower than significance level of 0.05 or 0.10, which means that these coefficients are statistically significant.

Table 4. First Stage of 2SLS estimation

LnP	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
District						
Carabayllo	0.706	0.091	7.760	**0.000**	0.527	0.884
Comas	-0.270	0.126	-2.150	**0.032**	-0.518	-0.023
Puente piedra	0.373	0.122	3.070	**0.002**	0.134	0.612
San Juan de Lurigancho	0.131	0.084	1.550	**0.122**	-0.035	0.296
San Juan de Miraflores	0.351	0.054	6.520	**0.000**	0.245	0.456
San Martin de Porres	-0.050	0.061	-0.820	**0.415**	-0.171	0.071
Villa el salvador	0.148	0.080	1.850	**0.065**	-0.009	0.304
Villa maría del triunfo	0.334	0.063	5.280	**0.000**	0.210	0.458
Lurín	0.463	0.106	4.380	**0.000**	0.255	0.670
Number_members	-0.007	0.008	-0.820	0.410	-0.023	0.009
Transportation_Time	0.001	0.002	0.810	0.416	-0.002	0.005
StorageCap	-0.142	0.037	-3.800	0.000	-0.215	-0.069
DummyIncome	0.089	0.042	2.120	0.034	0.007	0.172
Constant	2.511	0.070	35.940	0.000	2.373	2.648

For the exogeneity condition, the variable district should not be correlated with u, $cov(district, u) = 0$. I can mention that district does not share common causes with consumption of water. In this regard, the figure below shows, as reference, that there is no relationship between district and residuals of the second stage of 2SLS estimation from Table 6.

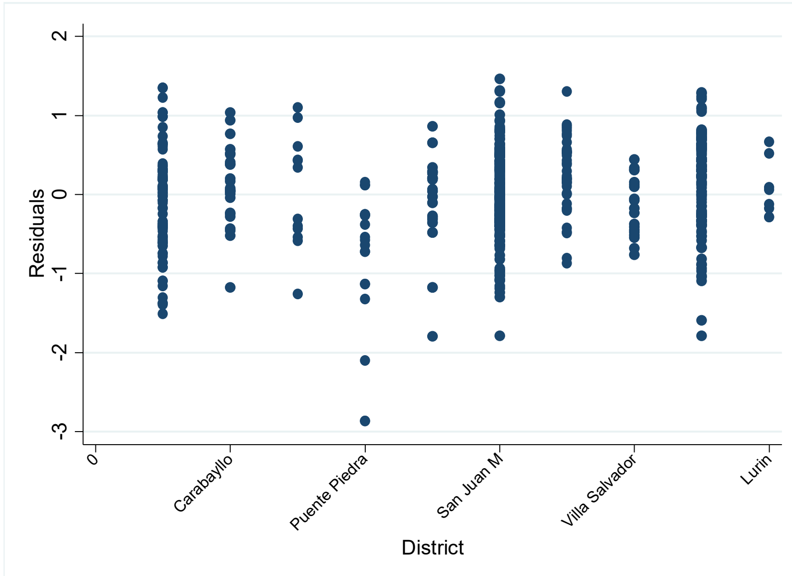

Figure 5. Relationship Between Residual and District

Finally, to detect weak instruments, I use the rule from Staiger and Stock (1997) on the size of F statistic because there are more than two instruments. Table 5 presents the results of the test for weak instruments.

Table 5. Test for weak instruments

F test of excluded instruments:	
F(9, 590)	**21.22**
Prob > F	0.00

Because the F statistic is larger than 10, I reject the null hypothesis that instruments are weak. In conclusion, the dummy variable for each district would be a valid, relevant and strong instrument.

4.1.2. Estimation of water demand and discussion

Table 6 presents the results of the second stage:

Table 6. Estimation of Water Demand by 2SLS estimation

LnQ	**Coef.**	**Std. Err.**	**t**	**P>t**	**[95% Conf. Interval]**	
LnP	-0.396	0.127	-3.110	0.002	-0.645	-0.146
Number_members	0.062	0.014	4.570	0.000	0.036	0.089
Transportation_Time	-0.002	0.003	-0.850	0.395	-0.008	0.003
StorageCap	0.527	0.053	9.970	0.000	0.423	0.630
DummyIncome	0.157	0.056	2.800	0.005	0.047	0.267
Constant	1.977	0.369	5.360	0.000	1.254	2.701

Note: A specification analysis about this model is given in Appendix C.

First, there is a negative relationship between consumption of water (demand) and the water price, which is consistent with Microeconomic theory. The price elasticity of demand was -0.396; if the water price increases by 1%, the consumption of water decreases by 0.396% which is statistically significant at the 1 percent level. The estimated demand is relatively inelastic, because the percentage change in consumption of water is less than the percentage change in water price. This was expected because water is a good which does not have many substitutes. Therefore, water price is statistically significant to explain variations of consumption of water.

Second, there is a positive relationship between water demand and number of people in household; if the number of people increases by one, the consumption of water increases by 6.43% and is statistically significant at the 1 percent level.[7]

Third, there is a negative relationship between water demand and transportation time from water purchase point to house, so if the transportation time increases by one minute, the consumption of water decreases by 0.24%. Nonetheless, the coefficient is not statistically significant different from zero.

[7]For estimation the semi-elasticity for these variables: Number_members, Transportation_Time, StorageCap and DummyIncome , we used: $\%\Delta\hat{Q} = 100\left(e^{\beta_i \Delta x_i} - 1\right)$

Fourth, there is a positive relationship between water demand and the storage capacity of water that each household has; if the storage capacity increases by one liter, the consumption of water increases by 69.35%, significant at the 1 percent level.

Fifth, there is a positive relationship between water demand and income, so if a household has income more than USD 315 monthly, the consumption of water is 17.01% more than a household with an income less than USD 315.

On the other hand, the following table compares, as reference, estimations using OLS, 2SLS and different variants of Generalized Method of Moment (GMM). We can notice that the estimated coefficients have changed by less 10% among 2SLS model and GMM models. However, OLS result about price coefficient has a difference bigger than 10% respect to 2SLS result, so it could be a signal that the instruments are working well.

Table 7. Comparison of models

LnQ	OLS	2SLS	GMM_het(1)	GMM_igmm(2)
LnP	-0.331***	-0.396**	-0.381**	-0.382**
Number_members	0.063***	0.062***	0.061***	0.058***
Transportation_Time	-0.003	-0.002	-0.003	-0.003
StorageCap	0.536***	0.527***	0.483***	0.477***
DummyIncome	0.151**	0.157**	0.236***	0.252***
Constant	1.798***	1.977***	1.963***	1.978***

Legend: * p<.05; ** p<.01; *** p<.001
(1) Optimal GMM given heteroskedastic errors. (2) GMM to iterate to converge.

4.2. Impacts on the social welfare

As mentioned previously in the methodology, I estimate the variation in social welfare as consequence of variation in water price, given that price is a determinant of the water demand that can be regulated in short run. In this regard, I estimate the variation in social welfare as the sum of variation in consumer welfare and producer welfare.

For measuring the consumer welfare, we have used three measurements: (i) variation in consumer surplus, (ii) compensating variation and (iii) equivalent variation.

Whereas for measuring the produce welfare, I use the variation in producer surplus. Table 8 shows the estimation result, simulating different levels of reducing of water price, which were obtained using equations, (22), (23) and (24):

Table 8. Estimation of Variation in Consumer and Producer Welfare

Objective of public policy on water price	Water price (PEN/m^3)	Compensating Variation (CV)	ΔConsumer Surplus (ΔCS)	Equivalent Variation (EV)	ΔProducer Surplus (ΔPS)
		(USD '000/Annual)			
Reduce by 50% (1)	7.89	26,420.67	27,344.88	29,523.38	-18,828.95
Reduce by 55% (1)	7.10	29,919.47	31,188.42	34,283.79	-21,623.58
Reduce by 60% (1)	6.31	33,484.28	35,200.67	39,587.00	-24,608.93
Reduce by 65% (1)	5.52	37,115.10	39,418.29	45,609.03	-27,821.66
Reduce by 70% (1)	4.73	40,811.94	43,885.31	52,621.29	-31,335.12
Reduce by 75% (1)	3.95	44,574.80	48,660.39	61,085.88	-35,244.68
Reduce by 80% (1)	3.16	48,403.67	53,824.23	71,794.98	-39,726.36
Price = AC (2)	2.35	52,379.24	59,670.23	97,636.19	-45,249.62

Note: (1) The reducing is respect to current water price (15.78 PEN/m^3) which is paid by consumers.

(2) PEN 2.35 is the average cost of water service established by SEDAPAL Tariff Study for the 2015-2020 regulatory period.

(3) Appendix D shows more details about the estimations.

Consumer surplus increases considerably because consumers consume more when the price is lower. Likewise, reducing of water price leads to have positive values of the compensating variation and the equivalent variation, which get bigger when a reduction of water price is larger. It is important to mention that the relation between variation in consumer surplus, compensating variation and equivalent variation is reasonable and consistent with equation (13).

On the other hand, the producer surplus decreases, because the profit of firms is reduced. However, for the society, the impact is positive since the variation in consumer surplus, compensating variation or equivalent variation are greater than variation in producer surplus, as presented in Table 9:

Table 9. Estimation of Variation in Social Welfare

Objective of public policy on water price	Water price (PEN/m^3)	Social Welfare (Using CV)	Social Welfare (Using ΔCS)	Social Welfare (Using EV)
			(USD '000/Annual)	
Reduce by 50%	7.89	7,591.73	8,515.94	10,694.43
Reduce by 55%	7.10	8,295.89	9,564.84	12,660.21
Reduce by 60%	6.31	8,875.35	10,591.74	14,978.07
Reduce by 65%	5.52	9,293.45	11,596.64	17,787.38
Reduce by 70%	4.73	9,476.82	12,550.19	21,286.17
Reduce by 75%	3.95	9,330.12	13,415.72	25,841.21
Reduce by 80%	3.16	8,677.30	14,097.87	32,068.62
Price = AC	2.35	7,129.62	14,420.61	52,386.57

As expected, there is a negative relationship between water price and social welfare regardless which measurement I use to estimate the consumer welfare. So, when the price is lower, social welfare will be greater. In an extreme case, if the government regulates the water price until it becomes similar to the average cost of water service provided by SEDAPAL, the social welfare would increase by 14,420.62 thousand of USD per year considering the variation of consumer surplus as consumer welfare, but if we consider the equivalent variation, the social welfare can increase up to 52,386.57 thousand of USD.

Since, a reduction in the water price in marginalized urban areas in Lima-Peru, where households do not have access to drinking water services, can significantly increase the social welfare, then, designing public policies to regulate the water price is strongly recommended.

In this regard, there could be two public policies which can help reduce the water price: (i) Economic regulation, where the government, through a regulator, define the water price level using regulation mechanism such as Rate of Return, which also guarantee the sustainability of water service in the long run; and (ii) increasing the coverage of water service provided by SEDAPAL. This would allow households who do not have access to potable water services with pipes to become consumers of SEDAPAL and get benefits such as lower price. This kind of policy produces several positives externalities, for example, better quality of water, decreased water transportation time, better continuity of service, etc.

V. CONCLUSION

This research reveals that consumption of water at marginalized urban areas in Lima-Peru, where households are not provided potable water service with pipes by SEDAPAL, depends significantly on water price. In particular, the price elasticity of demand equals -0.396 which means that the estimated demand is relatively inelastic in accordance with microeconomic theory. Likewise, explanatory variables such as household members, the storage capacity of water that each household has, and the level of income are statistically significant to explain variations in water demand. However, transportation time from water purchase point to house is not statistically significant. On the other hand, there is a strong negative relation between water price and social welfare, i.e., reduction in price leads to increasing of social welfare due to an increase in consumer surplus, as well as the compensating variation or the equivalent variation are greater than reduction in the producer surplus. Based in both results, designing public policies to regulate the level of water price are strongly recommended to increase the social welfare of marginalized urban sectors in Lima-Peru.

REFERENCES

Al-Qunaibet, M., and Jhonston, R. (1985). Municipal demand for water in kuwait: methodological issues and empirical results. *Water Resources Research*, *21*, (4), 433-438.

Billings, B., and Agthe, D. (1980). Price elasticities for water: A case of increasing block rates. *Land Economics*, *56*(1), 73-84.

Bonifaz, J. ., and Aragón, G. (2012). *Sobrecostos por la falta de infraestructura en agua potable: una aproximación empírica.* Lima: Universidad del Pacífico.

Cameron, C., and Trivedi, P. (2009). *Microeconometrics using Stata*. Place: Stata Press.

Cavanagh, S., Hanemann, M., and Stavins, R. (2001). Muffled price signals: household water demand under increasing-block prices (FEEM Working Paper No. 40.2002). Massachusets: Harvard University.

Chiang, A. (2005). *Fundamental methods of mathematical economics* (4th edition). New York: McGraw-Hill Irwin..

Directo Marketing Research and Consulting (2015). *Consultoría para la ejecución de encuesta sobre formas de acceso y gasto en agua potable en los distritos del sur de Lima y a usuarios en el ámbito de Sedapal sin acceso al servicio*. Lima: Autor.

Harberger, A. (1954). Monopoly and resource allocation. *American Economic Review*. *44*(2), 77-87.

Hausman, J. (1985). The Econometrics of nonlinear budget sets. *Econometrica*, *53*(6), 1255-1282.

Hewitt, J. and Hanemann, M. (1995). A Discrete/Continuous choice approach to residential water demand under block rate pricing. *Land Economics, 71*(2), 173-192.

Hicks, J. (1939). The foundations of welfare economics. *The Economic Journal*, *49*(196), 696-712

Huamani, S. (2017). *Estimación de la rentabilidad social de incrementar la cobertura de agua potable en Lima Metropolitana*. Lima: Universidad del Pacífico.

Hutton, G., Haller, L., and Bartram, J. (2007). Global cost-benefit analysis of water supply and sanitation interventions. *Journal of Water and Health*, *5*(4), 481-502.

Jaramillo-Mosqueira, L. (2005). Evaluación econométrica de la demanda de agua en uso residencial en México. *El Trimestre Económico, 72*(286), 367-390.

Jehle, F. and Reny, P. (2011). *Advanced Microeconomic Theory* (3rdEdition). Londong: Pearson Education Limited.

Jiménez, D., Orrego, S., Vásquez, F. and Ponce, R. (2017). Estimación de la demanda de agua para uso residencial urbano usando un modelo discreto-continuo y datos desagregado a nivel de hogar: El caso de la ciudad de Manizales, Colombia. *Lecturas de Economía,* (86), 153-178.

Kreps, D (2012). *Microeconomic Foundations I: Choice and Competitive Markets.* New Jersey: Princeton University Press.

Mas-Colell, A., Whinston, M. and Green, J. (1995). *Microeconomic Theory.* Londres: Oxford Economic Press.

Miyawaki, K., Yasuhiro, O. and Hibiki, A. (2011). Panel data analysis of Japanese residential water demand using a Discrete/Continuous choice approach. *The Japanese Economic Review*, *62*(3), 365-386.

Nauges, C. and Whittington, D. (2010). Estimation of water demand in developing countries: An overview. *World Bank Research Observer*, *25*(2), 263-294.

Nordin, J. (1976). A proposed modification of Taylor´s demand analysis: Comment. *The Bell Journal of Economics*. *7*(2), 719-21.

Oblitas, L. (2010). *Servicios de agua potable y saneamiento en el Perú: beneficios potenciales y determinantes de éxito.* Santiago de Chile: CEPAL.

Olmstead, S. (2009). Reduced-Form versus structural models of water demand under nonlinear prices. *Journal of Business and Economic Statistics*, *17*(1), 84-94-

Opaluch, J. (1982). Urban residential demand for water in the United States: Futher discussion. *Land Economics*, *58*, 225-27.

Ortiz, H. and Bendezu, L. (2006). *¿Cómo compatibilizar costos del servicio con el bienestar de los usuarios residenciales?: Un análisis del servicio de agua potable en el Perú*. Lima: CIES.

Pint, E. (1999). Household responses to increased water rates during the California drought. *Land Economics*, *75*, (2), 246-266.

Rawls, J. (1985). Justice as fairness: Political not metaphysical. *Philosophy and Public Affairs*, *14*(3), 223-25.

Sen, A. (1997). *On Economic Inequality*. Oxford: Clarendon Press.

Staiger, D. and Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557-586.

Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression.* New York: Cambrige University Press.

SUNASS (2015). *Estudio Tarifario: Determinación de la fórmula tarifaria, estructuras tarifarias y metas de gestión aplicables a la Empresa Servicio de Agua Potable y Alcantarillado de Lima*. Lima: Author.

Stiglitz, J. (2000). *Economics of the pucblic sector* (3rd Edition). New York: Columbia University Press.

Taylor, L. (1975). The demand for electricty: A survey. *The Bell Journal of Economics*. *6*(1), 74-110.

Varian, H. (1992). *Microeconomic Analysis* (3rd Edition). New York: *Library of Congress Cataloging in Publication Data.*

Wooldridge, J. (2015). *Introductory Econometrics: A Modern Approach.* (6th Edition). Boston: Cengage.

VII. APPENDIX

Appendix A

How does social welfare change when some determinants (different from the price) change?

A variation in x_2, with p fixed, also causes the consumer to choose a different market basket. Now, let us suppose that one explanatory variable is better off (in this case, increases income), it generates a shift to the right of the water demand curve, which is shown in the following figure.

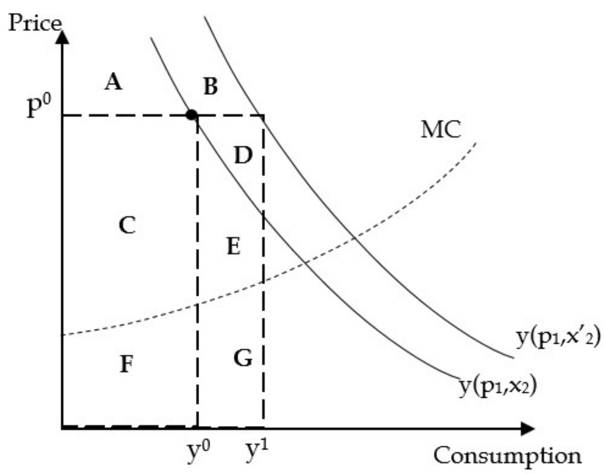

Figure 6. Variation in other Variables Different from Price

Similar to price variation analysis, it is necessary to identify the level of consumer surplus and producer surplus before and after variation in x_2. According to after-before analysis, the variation in social welfare is B+D+E area, as shown in the following table.

Table 10. Variation in social welfare (Before - After) by Income Variation

Variable	Before	After	Variation
Consumer surplus	A	A+B	B
Producer surplus	C	C+D+E	D+E
Social welfare	A+C	A+B+C+D+E	**B+D+E**

Mathematically, to calculate B+D+E area, it is necessary to apply the following equations. For variation of consumer surplus:

$$B = \int_{p0}^{p'max} y(p, x'_2)dp \; - \int_{p0}^{p\,max} y(p, x_2)dp \qquad (26)$$

For variation of producer surplus:

$$D + E = p^0(y^1 - y^0) - \int_{y^0}^{y^1} MC(y)\, d_y \qquad (27)$$

$$D + E = p^0(y^1 - y^0) - AC(y^1 - y^0) = (p^0 - AC)(y^1 - y^0)$$

Finally, the variation in social welfare in this case (Δx_2) is defined by:

$$\Delta W = \int_{p0}^{p'max} y(p, x'_2)dp \; - \int_{p0}^{p\,max} y(p, x_2)dp \; + (p^0 - AC)(y^1 - y^0) \qquad (28)$$

Appendix B

Derivation of the Hicksian Demand of figure 3

We can obtain y_h^{1CV} and y_h^{1EV} using the following equations:

$$y_h^{1CV} = y^0 - \frac{\partial y_h}{\partial p}(p^0 - p^1) \tag{29}$$

$$y_h^{1EV} = y^1 + \frac{\partial y_h}{\partial p}(p^0 - p^1) \tag{30}$$

Where $\frac{\partial y_h}{\partial p}$ is the slope of the Hicksian demand curve. To calculate it, we will use Slutsky equation which relates changes in Marshallian demand to changes in Hicksian demand, which is known as such since it compensates to maintain a fixed level of utility. Formally, the Slutsky equation is defined by:

$$\frac{\partial y}{\partial p} = \frac{\partial y_h}{\partial p} - y\frac{\partial y}{\partial I} \tag{31}$$

The slope of Marshallian demand curve, $\frac{\partial y}{\partial p}$, can be calculated using the estimated price elasticity which is shown in the table 6, the initial value of water price and initial value of consumption of water. However, the derivation of demand respect to income, $\frac{\partial y}{\partial I}$, cannot be taken from table 6 because the income variable is a dummy. So, we will take the value of this derivation from Ortiz and Bendezu (2006) which has a reasonable value equal to 0.0013.

Appendix C

Specification analysis

This appendix presents an evaluation of the econometrics model showed in table 6. The fitted model has $R^2 = 0.226$, which is reasonable for cross-section data. Likewise, four out of five regressors are highly statistically significant with the expected coefficient signs.

On the other hand, in the next figure, it can be observed that the average of residuals would be around zero, with a dispersion quite concentrated in this average. We can infer from this that there is homoscedasticity, so it would be expected that the estimated coefficients are efficient.

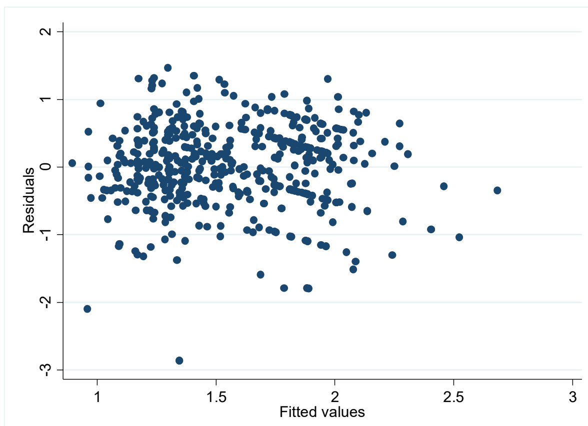

Figure 7. Residual Plot

Likewise, applying the Breusch-Pagan test for heteroscedasticity, we fail to reject the null hypothesis about constant variance because of the p-value (0.8245), so we also can infer that there is homoscedasticity.

Table 11. Breusch-Pagan Test for heteroskedasticity

Variable	Value
Chi2(1)	0.05
Prob > chi2	0.8245

On the other hand, in the following figure, it can be observed that the average of residuals would be around zero, with a dispersion quite concentrated in this average. We can infer from this that there is homoscedasticity, so it would be expected that the estimated coefficients are efficient.

Then, histogram of the residuals and Kernel density estimated, which are shown in the following figures, reveal that the distribution of the residuals look like a normal distribution, so we would say that the model has a reliable inference.

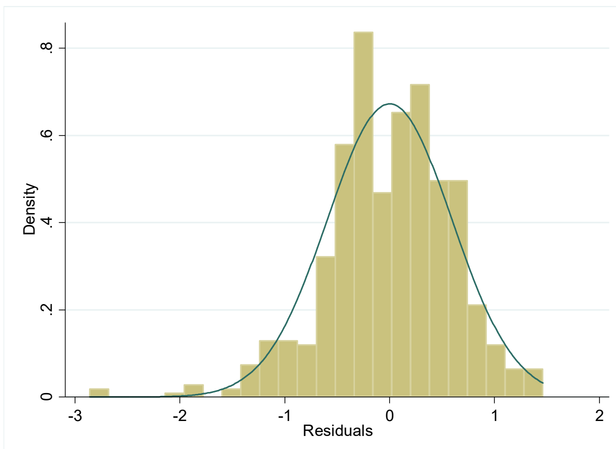

Figure 8. Histogram of Residuals

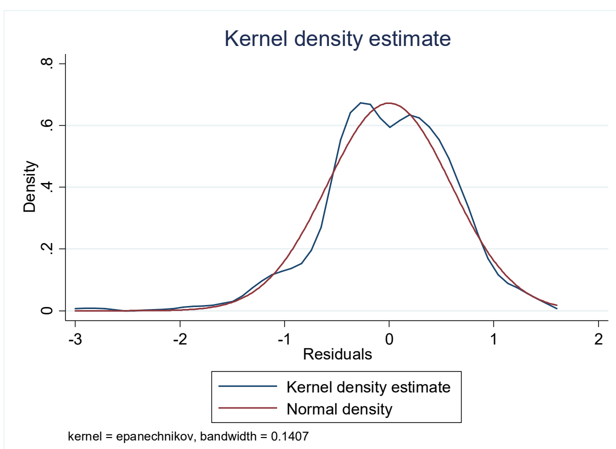

Figure 9. Kernel Density Estimation

Finally, about multicollinearity test, the mean of variance inflation factor (VIF) is less than 10, which means that there would be no relationship between the explanatory variables, so a multicollinearity problem would be discarded.

Table 12. Variance Inflation Factor

Variable	**VIF**	**1/VIF**
LnP	1.27	0.787
Number_members	1.11	0.898
Transportation_Time	1.08	0.923
StorageCap	1.16	0.862
DummyIncome	1.15	0.869
Mean VIF	1.16	

Appendix D

Estimation of variation in social welfare because of reducing in water price by different cases

Reduce by 50%	**Average household (PEN) Monthly**	**Total of households (PEN '000) Monthly**	**Total of households (USD '000) Monthly**	**Total of households (USD '000) Annual**
$\Delta CS = B + C$	37.28	7,291.97	2,278.74	**27,344.88**
$B = (p^0 - p^1)y^0$	33.06	6,466.54	2,020.79	24,249.51
C = (B + C)-B	4.22	825.43	257.95	3,095.37
$E = (p^1 - AC)(y^1 - y^0)$	7.39	1,445.48	451.71	5,420.57
$\Delta PS = E - B$	-25.67	-5,021.05	-1,569.08	**-18,828.95**
$CV = B + C_1$	36.02	7,045.51	2,201.72	**26,420.67**
$EV = B + C'_1 + C2$	40.25	7,872.90	2,460.28	**29,523.38**
$\Delta W_1 = \Delta CP + \Delta SP$	11.61	2,270.92	709.66	**8,515.94**
$\Delta W_2 = CV + \Delta SP$	10.35	2,024.46	632.64	**7,591.73**
$\Delta W_3 = EV + \Delta SP$	14.58	2,851.85	891.20	**10,694.43**

Reduce by 55%	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	42.52	8,316.91	2,599.04	**31,188.42**
$B = (p^0 - p^1)y^0$	37.13	7,262.63	2,269.57	27,234.86
C = (B + C)-B	5.39	1,054.28	329.46	3,953.57
$E = (p^1 - AC)(y^1 - y^0)$	7.65	1,496.34	467.61	5,611.28
ΔPS= E – B	-29.48	-5,766.29	-1,801.97	**-21,623.58**
CV = B + C_1	40.79	7,978.52	2,493.29	**29,919.47**
EV = B + C'_1 + C2	46.74	9,142.34	2,856.98	**34,283.79**
ΔW_1= ΔCP + ΔSP	13.04	2,550.62	797.07	**9,564.84**
ΔW_2= CV + ΔSP	11.31	2,212.24	691.32	**8,295.89**
ΔW_3= EV + ΔSP	17.26	3,376.06	1,055.02	**12,660.21**

Reduce by 60%	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	47.99	9,386.84	2,933.39	**35,200.67**
$B = (p^0 - p^1)y^0$	41.21	8,060.68	2,518.96	30,227.54
C = (B + C)-B	6.78	1,326.17	414.43	4,973.13
$E = (p^1 - AC)(y^1 - y^0)$	7.66	1,498.30	468.22	5,618.61
ΔPS= E – B	-33.55	-6,562.38	-2,050.74	**-24,608.93**
CV = B + C_1	45.65	8,929.14	2,790.36	**33,484.28**
EV = B + C'_1 + C2	53.97	10,556.53	3,298.92	**39,587.00**
ΔW_1= ΔCP + ΔSP	14.44	2,824.46	882.65	**10,591.74**
ΔW_2= CV + ΔSP	12.1	2,366.76	739.61	**8,875.35**
ΔW_3= EV + ΔSP	20.42	3,994.15	1,248.17	**14,978.07**

Reduce by 65%	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	53.74	10,511.54	3,284.86	**39,418.29**
$B = (p^0 - p^1)y^0$	45.29	8,858.72	2,768.35	33,220.22
C = (B + C)-B	8.45	1,652.82	516.51	6,198.08
$E = (p^1 - AC)(y^1 - y^0)$	7.36	1,439.62	449.88	5,398.56
ΔPS= E – B	-37.93	-7,419.11	-2,318.47	**-27,821.66**
CV = B + C_1	50.6	9,897.36	3,092.93	**37,115.10**
EV = B + C'_1 + C2	62.18	12,162.41	3,800.75	**45,609.03**
ΔW_1= ΔCP + ΔSP	15.81	3,092.44	966.39	**11,596.64**
ΔW_2= CV + ΔSP	12.67	2,478.25	774.45	**9,293.45**
ΔW_3= EV + ΔSP	24.25	4,743.30	1,482.28	**17,787.38**

Reduce by 70%	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	59.83	11,702.75	3,657.11	**43,885.31**
$B = (p^0 - p^1)y^0$	49.37	9,656.77	3,017.74	36,212.90
C = (B + C)-B	10.46	2,045.98	639.37	7,672.41
$E = (p^1 - AC)(y^1 - y^0)$	6.65	1,300.74	406.48	4,877.78
ΔPS= E – B	-42.72	-8,356.03	-2,611.26	**-31,335.12**
CV = B + C_1	55.64	10,883.18	3,401.00	**40,811.94**
EV = B + C'_1 + C2	71.74	14,032.34	4,385.11	**52,621.29**
ΔW_1= ΔCP + ΔSP	17.11	3,346.72	1,045.85	**12,550.19**
ΔW_2= CV + ΔSP	12.92	2,527.15	789.74	**9,476.82**
ΔW_3= EV + ΔSP	29.02	5,676.31	1,773.85	**21,286.17**

Reduce by 75%	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	66.34	12,976.10	4,055.03	**48,660.39**
$B = (p^0 - p^1)y^0$	53.45	10,454.82	3,267.13	39,205.58
C = (B + C)-B	12.89	2,521.28	787.90	9,454.82
$E = (p^1 - AC)(y^1 - y^0)$	5.4	1,056.24	330.08	3,960.90
ΔPS= E – B	-48.05	-9,398.58	-2,937.06	**-35,244.68**
CV = B + C_1	60.77	11,886.61	3,714.57	**44,574.80**
EV = B + C'_1 + C2	83.28	16,289.57	5,090.49	**61,085.88**
ΔW_1= ΔCP + ΔSP	18.29	3,577.52	1,117.98	**13,415.72**
ΔW_2= CV + ΔSP	12.72	2,488.03	777.51	**9,330.12**
ΔW_3= EV + ΔSP	35.23	6,890.99	2,153.43	**25,841.21**

Reduce by 80%	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	73.38	14,353.13	4,485.35	**53,824.23**
$B = (p^0 - p^1)y^0$	57.53	11,252.87	3,516.52	42,198.26
C = (B + C)-B	15.85	3,100.26	968.83	11,625.98
$E = (p^1 - AC)(y^1 - y^0)$	3.37	659.17	205.99	2,471.90
ΔPS= E – B	-54.16	-10,593.70	-3,310.53	**-39,726.36**
CV = B + C_1	65.99	12,907.64	4,033.64	**48,403.67**
EV = B + C'_1 + C2	97.88	19,145.33	5,982.92	**71,794.98**
ΔW_1= ΔCP + ΔSP	19.22	3,759.43	1,174.82	**14,097.87**
ΔW_2= CV + ΔSP	11.83	2,313.95	723.11	**8,677.30**
ΔW_3= EV + ΔSP	43.72	8,551.63	2,672.39	**32,068.62**

Price = AC	Average household (PEN) Monthly	Total of households (PEN '000) Monthly	Total of households (USD '000) Monthly	Total of households (USD '000) Annual
ΔCS = B + C	81.35	15,912.06	4,972.52	**59,670.23**
$B = (p^0 - p^1)y^0$	61.69	12,066.56	3,770.80	45,249.62
C = (B + C)-B	19.66	3,845.50	1,201.72	14,420.61
$E = (p^1 - AC)(y^1 - y^0)$	0	0	0	0
ΔPS= E – B	-61.69	-12,066.56	-3,770.80	**-45,249.62**
CV = B + C_1	71.41	13,967.80	4,364.94	**52,379.24**
EV = B + C'_1 + C2	133.11	26,036.32	8,136.35	**97,636.19**
ΔW_1= ΔCP + ΔSP	19.66	3,845.50	1,201.72	**14,420.61**
ΔW_2= CV + ΔSP	9.72	1,901.23	594.14	**7,129.62**
ΔW_3= EV + ΔSP	71.42	13,969.75	4,365.55	**52,386.57**

Note: Exchange rate used: 3.20 (PEN/USD)

Note: For B + C, we have used this equation: $\int_{p1}^{p0} y(p, x_2)dp = e^{(\beta_0 + \sum_{i\neq 1}\beta_i x_i)}\left(\frac{p_0^{\beta_1+1} - p_1^{\beta_1+1}}{\beta_1+1}\right)$